Popular Agriculture 2(5) 143-146(September, 2026)
Published at popularagriculture.com

Global E-Magazine
ISSN: 3139-7603

POPULAR ARTICLE

Smart breeding with genomic selection and NGS from genome data to improved crop varieties

T. Vijay[1], M.S. Hanif[2], N. Asif[3], B. Srikanth[4], R. G. Roy[5]

1 Department of Genetics and Plant Breeding, Junagadh Agricultural University, Junagadh 362001, Gujarat, India

2 Centre for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang 050021, China

3 College of Fisheries and Life Sciences, Shanghai Ocean University, Shanghai 201306, China

4 Department of Genetics and Plant Breeding, Mahatma Phule Krishi Vidyapeeth, Pune 411005, Maharashtra, India

5 Department of Genetics and Plant Breeding, Tantia University, Sri Ganganagar 335002, Rajasthan, India

Corresponding authors email: vijayraajputthakur51297@gmail.com

Manuscript received: August 4, 2026; Decision on manuscript, August 5, 2026; Manuscript accepted: September 11, 2026

Introduction

For most of the last century, plant breeding was a numbers game played out in the field. A breeder crossed two promising parents, grew out thousands of progeny, and waited often three to five years to see which plants flowered earlier, yielded more, or shrugged off disease. Selection depended on what the eye could see and the yardstick could measure. It worked, but it was slow, and it could only act on traits that had already shown up in the plant. Genomic selection (GS), powered by next-generation sequencing (NGS), has changed that equation. Instead of waiting for a trait to appear in the field, breeders can now read a seedling's DNA and predict, with reasonable accuracy, how it will perform as an adult plant before a single flower opens. First proposed as a statistical concept by Meuwissen, Hayes and Goddard in 2001, the idea sat largely on the shelf for years because genotyping thousands of individuals at thousands of markers was too expensive. The arrival of affordable, high-throughput sequencing changed genomic selection has since become one of the most consequential shifts in breeding methodology since the rediscovery of Mendel's laws.

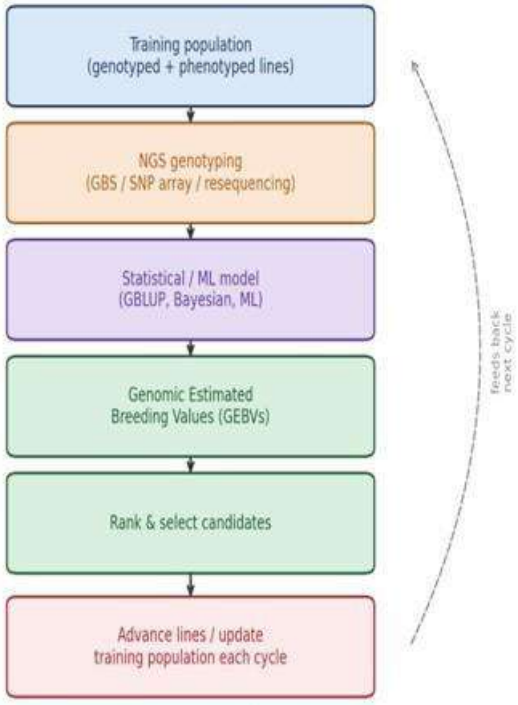
What genomic selection actually does?

Genomic selection is often confused with marker-assisted selection (MAS), but the two work quite differently. MAS tracks a handful of markers tightly linked to one or two major genes useful for a simple trait like a single disease-resistance gene, but blind to traits controlled by hundreds of small-effect genes scattered across the genome, such as yield, drought tolerance, or grain quality. The practical gain is speed. A controversial

Popular Agriculture 2(5) 143-146(September, 2026)
Published at popularagriculture.com
Global E-Magazine
ISSN: 3139-7603

breeding cycle for a self -pollinated crop can take 5 to 7 years from the initial cross to variety release. Genomic selection compresses that considerably by letting breeders cycle through generations on genotype alone, reserving expensive, time-consuming field phenotyping mainly for late-stage confirmation (Crossa *et al.*, 2017).

Fig 1: A geneomic selection procedure



Why NGS was the missing piece

Genomic selection stayed a theoretical curiosity for the better part of a decade after Meuwissen and colleagues first described it, mainly because genotyping thousands of lines at thousands of markers was prohibitively expensive at the time (Kumar *et al.*, in Genomic Selection in the Era of Next-Generation Sequencing, 2016). Next-generation sequencing changed the economics of the whole enterprise. A handful of NGS-based genotyping approaches now dominate plant breeding pipelines: Genotyping-by-sequencing (GBS) uses restriction enzymes to reduce genome complexity before sequencing, generating tens of thousands of SNPs across a population at a fraction of the cost of whole-genome sequencing. SNP arrays offer fixed, highly consistent panels of pre-selected markers, well suited to large, established breeding programs. Whole-genome resequencing, increasingly affordable, captures not just SNPs but structural variation, giving the fullest picture of genetic diversity available. And skim sequencing combined with computational imputation offers a middle path lower cost, with dense marker information recovered statistically rather than measured directly. Whichever platform is chosen, the output feeding into the breeding pipeline is the same: a large matrix of lines by markers, ready to be handed to a genomic prediction model.

Popular Agriculture 2(5) 143-146(September, 2026)
Published at popularagriculture.com
Global E-Magazine
ISSN: 3139-7603

Building and using the prediction model

Turning marker data into breeding decision requires statistical machinery, and a good deal of methodological research has gone into refining it since 2001.GBLUP (genomic best linear unbiased prediction) treats the combined effect of all markers as a single genomic relationship matrix, and remains a workhorse in many programs for its robustness and relative simplicity. Bayesian methods such as BayesA, B and Cπ allow marker effects to vary in size, which can better capture traits shaped by a few genes of larger effect alongside many of smaller effect (Habier *et al.*, 2011). More recently, machine learning approaches random forests, support vector machines, and various deep learning architectures have been explored for traits with complex, non-additive genetic architecture, though they typically need larger training populations to reliably outperform the classical statistical models (Montesinos-López *et al.*, 2022). Model accuracy depends heavily on the size and genetic relatedness of the training population, the heritability of the trait in question, and how well the training population represents the population it is being used to predict. A model trained on temperate maize inbreds, for instance, will not predict tropical germplasm well; genetic distance between the training and prediction sets is one of the most common reasons genomic selection (Voss-Fels *et al.*, 2019).

A practical workflow, cycle after cycle

In a modern breeding program, a genomic selection cycle typically runs as follows. A training population is assembled with both genotypic and multi-location phenotypic data. That data trains a prediction model, which is cross-validated to estimate its accuracy before it is trusted with real decisions. The candidate population new crosses or breeding lines under consideration is then genotyped on the same marker platform, and GEBVs are calculated and ranked for every candidate. Top-ranked individuals are advanced, intercrossed, or retained as parents for the next cycle, often skipping a full phenotyping cycle entirely. Periodically, the training population is refreshed with newlyphenotyped lines, since prediction accuracy decays over generations as allele frequencies and patterns of linkage disequilibrium shift (Fu *et al.*, 2022). That last step matters more than it might appear. Genomic selection is not a one-time investment; it is a living system that needs fresh phenotypic data flowing back in, or its predictive power quietly erodes cycle.

Where it is already paying off ?

Genomic selection has moved well past the pilot-project stage in several major crops. Maize breeding programs, both public and private, use it routinely to advance doubled-haploid lines, and genomic selection applied to maize hybrid breeding has already shown tangible gains in genetic progress (Crossa *et al.*, 2017; Xu *et al.*, 2020). Wheat breeding networks, including CIMMYT's global programs, have integrated genomic selection fortraits such as yield potential and rust resistance, where the underlying genetics are too complex for simple marker-assisted approaches (Varshney *et al.*, 2021). Dairy cattle breeding pioneered much of the underlying statistical framework decades earlier, and plant breeders have been adapting those lessons ever since an interesting cross-pollination between animal and plant science that runs through most of the genomic selection literature. For orphan and minor crops including many ornamental and industrial species that never attracted the R&D budgets of maize or rice the falling cost of NGS is finally making genomic selection feasible where it was once unthinkable.

Popular Agriculture 2(5) 143-146(September, 2026)
Published at popularagriculture.com

Global E-Magazine
ISSN: 3139-7603

The fine print: limitations worth keeping in mind

Genomic selection is powerful, but it is not a silver bullet. It is only as good as the training population behind it poor-quality phenotypic data produces poor predictions no matter how sophisticated the statistical model. Genotype-by-environment interaction remains a persistent challenge; a GEBV calculated from trials in one set of environments may not transfer cleanly to a very different target environment.It does not eliminate field trials altogether final candidate varieties still need multi-location, multi-year testing before release, for regulatory reasons as much as practical ones. And the upfront investment is real: building a good training population and the bioinformatics infrastructure to support it takes years of sustained effort, which smaller programs may struggle to justify without institutional backing.

Looking ahead

The next frontier for genomic selection lies in combining it with other NGS-enabled tools: pangenome references that capture structural variation invisible to standard SNP arrays, multi-omics data layered on top of genomic predictions, and machine learning models trained on ever larger and more diverse datasets. As sequencing costs continue to fall and computational tools mature, genomic selection is likely to shift from a competitive advantage held by a few well-resourced programs to a standard tool available across public breeding institutions worldwide a genuinely democratizing shift for crop improvement. For a discipline that has always balanced patience with precision, genomic selection offers something breeders have wanted since Mendel: a way to see the genotype before the plant has even finished growing.

References

1. Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J. and Varshney, R.K. 2017. Genomic selection in plant breeding: Methods, models, and perspectives. Trends Plant Sci., 22(11): 961–975.
2. Habier, D., Fernando, R.L., Kizilkaya, K. and Garrick, D.J. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics, 12: 186. doi:10.1186/1471-2105-12-186.
3. Bhat, J.A., Ali, S., Salgotra, R.K., Mir, Z.A., Dutta, S., Jadon, V., Tyagi, A., Mushtaq, M.M., Jain, N., Singh, P.K., Singh, G.P. and Prabhu, K.V. et al. 2016. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Front. Genet., 7: 221. doi:10.3389/fgene.2016.00221.
4. Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4): 1819–1829. doi:10.1093/genetics/157.4.1819.
5. Montesinos-López, O.A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J.A., Martini, J.W.R., Fajardo-Flores, S.B. et al. 2021. A review of deep learning applications for genomic selection. BMC Genomics, 22: 19. doi:10.1186/s12864-020-07319-x